

PERBANDINGAN PENGGUNAAN ALGORITMA KRZYZAK DENGAN ALGORITMA RAMBATAN BALIK PIAWAI DALAM DOMAIN PERAMALAN

RAZANA ALWEE

ROSELINA SALLEHUDDIN

SITI MARIYAM HJ. SHAMSUDDIN

*Fakulti Sains Komputer & Sistem Maklumat
Universiti Teknologi Malaysia*

ABSTRAK

Artikel ini bertujuan mengkaji prestasi rangkaian neural menggunakan algoritma Krzyzak berbanding dengan algoritma rambatan balik piawai dalam domain peramalan. Kajian dilakukan terhadap data siri masa tak bermusim iaitu pengeluaran hasil kayu balak negara. Ukuran prestasi adalah berdasarkan dua perkara iaitu ketepatan ramalan yang dihasilkan dengan menggunakan ralat purata kuasadua (RMS) dan masa pembelajaran, iaitu masa yang diambil oleh setiap model untuk menumpu. Hasil kajian menunjukkan dengan pemilihan pemalar pembelajaran yang kecil, algoritma Krzyzak adalah lebih baik berbanding algoritma rambatan balik piawai bagi peramalan jangka sederhana dan jangka panjang.

Kata kunci: Rangkaian Neural, Algoritma Rambatan Balik, Algoritma Krzyzak, Peramalan, Data Siri Masa.

ABSTRACT

The purpose of this study is to compare the performance of neural network using Krzyzak algorithm and standard back propagation algorithm in forecasting domain. To implement this study a timber data set, which represents a non-seasonal time series data, is used. The performance is measured based on the accuracies, which is, quantified by root mean square error and learning speed for convergence. The results show that by using a small value of learning rate, Krzyzak algorithm is better than standard back propagation algorithm for medium and long term forecasting.

Keywords: Neural Network, Back Propagation Algorithm, Krzyzak Algorithm, Forecasting, Time Series Data.

PENGENALAN

Rangkaian neural merupakan suatu sistem pemprosesan maklumat yang mempunyai ciri-ciri perlakuan tertentu yang sama dengan sistem neural biologi (Fausett, 1994). Rangkaian neural terdiri daripada sejumlah elemen pemprosesan mudah yang dikenali sebagai neuron. Setiap neuron dihubungkan dengan neuron yang lain secara terus dengan pemberat yang tertentu. Nilai pemberat diperoleh melalui operasi pembelajaran yang dilakukan oleh rangkaian. Pemberat ini merupakan maklumat yang akan digunakan oleh rangkaian bagi menyelesaikan masalah dalam pelbagai bidang seperti klasifikasi pola, kelompok, penghampiran, pengoptimuman, kawalan dan peramalan.

Algoritma rambatan balik merupakan satu daripada operasi pembelajaran yang paling popular dalam rangkaian neural. Ini adalah kerana algoritma ini dapat meminimumkan ralat perbezaan di antara vektor output sebenar dan vektor output sasaran (Siti Mariyam Hj. Shamsuddin, Norazah Yusof, Paridah Samsuri, Nor Bahiah Ahmad & Siti Zaiton Mohd. Hashim, 2000). Algoritma ini telah berjaya digunakan dalam pelbagai bidang termasuklah peramalan data siri masa. Namun begitu, terdapat beberapa masalah dalam algoritma rambatan balik seperti masalah minimum setempat yang menyebabkan rangkaian kehilangan kebolehpayaan untuk belajar. Oleh yang sedemikian, Krzyzak (Krzyzak & Shen, 1990) telah mencadangkan pengubahsuaian fungsi kos dan berjaya digunakan dalam masalah pengecaman tulisan. Algoritma ini menumpu dengan lebih cepat dan mempunyai kadar pengecaman yang lebih baik berbanding algoritma rambatan balik (Krzyzak & Shen, 1990). Dalam kajian ini, prestasi algoritma Krzyzak berbanding algoritma rambatan balik piawai dalam peramalan data siri masa akan dibincangkan. Ketepatan ramalan dan kepantasan proses pembelajaran adalah dua aspek penting dalam peramalan data siri masa. Oleh itu tumpuan terhadap dua aspek ini telah dilakukan dalam membandingkan prestasi algoritma rambatan balik dan algoritma Krzyzak (Krzyzak & Shen, 1990).

RANGKAIAN NEURAL

Rangkaian neural merupakan satu kaedah pemprosesan data yang diilhamkan daripada biologi manusia. Rangkaian neural boleh dikategorikan berdasarkan kepada tiga perkara iaitu seni bina rangkaian, algoritma pembelajaran dan fungsi keaktifan. Rangkaian neural boleh terdiri daripada satu aras atau lebih. Jika rangkaian neural hanya mempunyai aras input dan aras output sahaja ianya dikenali

sebagai satu aras. Tetapi jika terdapat satu atau lebih aras tersembunyi di antara aras input dan aras output ianya dikenali sebagai multiaras. Rambatan balik dan Perseptron masing-masing merupakan contoh bagi seni bina rangkaian multiaras dan satu aras.

Pembelajaran rangkaian neural merujuk kepada proses mengemaskinikan nilai pemberat pada setiap talian yang terdapat pada setiap aras. Terdapat tiga jenis pembelajaran yang boleh digunakan iaitu pembelajaran tanpa selia, berselia dan rangsangan. Pembelajaran dikatakan tanpa selia jika penyelarasan pemberat pada talian dilakukan tanpa memerlukan perbandingan dengan output sasaran. Contoh pembelajaran tanpa selia ialah rangkaian Kohonen. Pembelajaran itu dikatakan pembelajaran berselia jika ianya melibatkan pemberian polar input satu persatu kepada rangkaian neural. Disamping itu, pembelajaran ini juga diberi output sebenar bagi tujuan perbandingan dengan output. Tambahan pula, ianya juga memerlukan set data latihan dan set data pengesahan. Contoh pembelajaran berselia ialah rangkaian rambatan balik.

Pembelajaran rangsangan pula merupakan pembelajaran yang terletak di antara pembelajaran selia dan tanpa selia. Ianya hanya memerlukan input dan gred atau isyarat ganjaran sebagai output. Namun begitu, kaedah ini kurang popular kerana ianya memerlukan masa yang lama untuk melatih rangkaian dan ianya lebih kompleks. Contoh pembelajaran rangsangan ialah pembelajaran automata. Fungsi keaktifan digunakan untuk menentukan hubungan yang wujud antara nod input dan nod output di dalam rangkaian. Perhubungan yang wujud mungkin merupakan hubungan linear atau tidak linear dan ianya bergantung kepada fungsi keaktifan yang dipilih. Antara fungsi keaktifan yang boleh digunakan ialah fungsi Sigmod, fungsi Identiti, fungsi Langkah, fungsi Tangen, fungsi Sinus dan fungsi Kosinus. Fungsi Identiti merupakan contoh fungsi linear manakala fungsi Langkah, fungsi Tangen, fungsi Sinus dan fungsi Kosinus merupakan contoh fungsi tidak linear. Pemilihan fungsi yang digunakan bergantung kepada masalah yang hendak diselesaikan.

Algoritma Rambatan Balik

Algoritma rambatan balik melibatkan dua fasa utama (Timothy, 1993) iaitu rambatan ke depan dan rambatan balik. Fungsi keaktifan yang digunakan ialah fungsi sigmoid. Dalam rambatan ke depan (Rajah 1), perhitungan O_j iaitu output unit j , ialah,

$$O_j = f(net_j) = 1 / (1 + e^{-net_j}) \quad (1)$$

$$net_j = \sum_i W_{ji} O_i + \theta_j$$

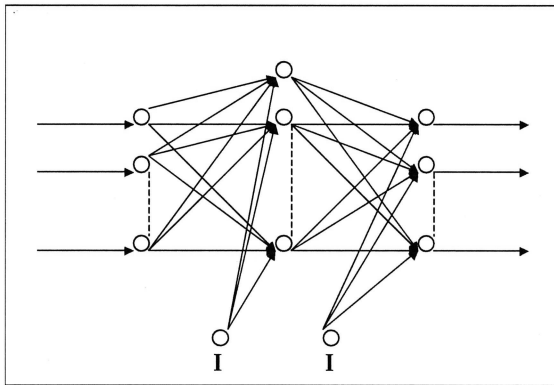
dengan W_{ji} adalah pemberat yang menghubungkan antara unit i ke unit j , O_i adalah output unit i dan θ_j adalah pemboleh ubah *bias*.

Dengan cara yang sama, perhitungan O_k iaitu output unit k , ialah,

$$O_k = f(net_k) = 1 / (1 + e^{-net_k}) \quad (2)$$

$$net_k = \sum_j W_{kj} O_j + \theta_k$$

dengan W_{kj} adalah pemberat yang menghubungkan antara unit j ke unit k , O_j adalah output unit j dan θ_k adalah pemboleh ubah *bias*.



Rajah 1
Rangkaian Suap ke Depan Tiga Lapisan

Tujuan pembelajaran algoritma ini adalah untuk meminimalkan nilai ralat kuasa dua yang dihasilkan oleh rangkaian (Roselina Sallehuddin, Mohd. Salihin Ngadiman & Siti Mariyam Hj. Shamsuddin, 2001). Ralat Min Kuasa Dua dinyatakan sebagai,

$$E = 1/2 (t_k - O_k)^2 \quad (3)$$

dengan O_k adalah nilai output dan t_k adalah nilai sasaran.

Dalam rambatan balik, pengemaskinian pemberat antara aras tersembunyi, j dan aras output, k dilakukan iaitu,

$$\Delta W_{kj}(t+1) = \eta \delta_k O_j + \alpha \Delta W_{kj}(t)$$

$$W_{kj}(t+1) = W_{kj}(t) + \Delta W_{kj}(t+1)$$

dengan

$$\delta_k = O_k (1 - O_k)(t_k - O_k) \quad (4)$$

Antara aras input, i dan aras tersembunyi, j

$$\Delta W_{ji}(t+1) = \eta \delta_j O_i + \alpha \Delta W_{ji}(t)$$

$$W_{ji}(t+1) = W_{ji}(t) + \Delta W_{ji}(t+1)$$

dengan

$$\delta_j = O_j (1 - O_j) \sum_k \delta_k \Delta W_{kj} \quad (5)$$

Masalah yang dihadapi dalam algoritma ini, seperti yang dibicarakan oleh Krzyzak (Krzyzak & Shen, 1990) adalah, $\delta_k = (t_k - O_k)(1 - O_k)O_k$ boleh menjadi sifar bukan sahaja apabila $t_k = O_k$ tetapi juga semasa $O_k = 0$ atau $O_k = 1$. Hal ini akan menyebabkan $\delta_k = 0$ juga. Oleh yang demikian, seluruh perkembangan kiraan akan menjadi sifar, dan rangkaian kehilangan kebolehpayaan untuk belajar.

Algoritma Krzyzak

Algoritma Krzyzak merupakan algoritma pembelajaran hasil daripada pengubahsuaian yang dilakukan terhadap algoritma rambatan balik piawai (Krzyzak & Shen, 1990). Apabila satu pola latihan dimasukkan ke dalam input rangkaian, output yang bersesuaian mesti tinggi atau output lain mesti rendah. Diberi satu output, O_k , dan nilai sasaran, $t_k = 0$, output akan memberi nilai berdasarkan fungsi kos seperti berikut,

$$- (1 - t_k) \ln (1 - O_k) \quad (6)$$

Jika nilai sasaran $t_k = 1$, fungsi yang digunakan ialah,

$$- t_k \ln O_k$$

dengan $1 \geq O_k \geq 0$. Bagi setiap pola p ditakrifkan fungsi seperti berikut:

$$E_p = - \sum_k (1 - t_{pk}) \ln (1 - O_{pk}) + t_{pk} \ln O_{pk} \quad (7)$$

Fungsi di atas boleh digunakan untuk menilai setiap output. Sekarang persamaan output akan menjadi,

$$\frac{\partial E_p}{\partial O_{pk}} = \frac{\partial}{\partial O_{pk}} \sum_k - (t_{pk} \ln O_{pk} + (1 - t_{pk}) \ln (1 - O_{pk}))$$

$$= t_{pk} \frac{1}{O_{pk}} + (1 - t_{pk}) \frac{1}{1 - O_{pk}}$$

Bagi nod output, rumus di atas boleh digantikan dengan,

$$\begin{aligned} \delta_{pk} &= - \left(t_{pk} \frac{1}{O_{pk}} + (1 - t_{pk}) \frac{1}{1 - O_{pk}} \right) (1 - O_{pk}) O_{pk} \\ &= (t_{pk} - O_{pk}) \end{aligned} \quad (8)$$

Oleh yang sedemikian, fungsi ralat di atas dikenali sebagai fungsi Krzyzak (Krzyzak & Shen, 1990). Persamaan bagi nod tersembunyi adalah sama seperti algoritma rambatan balik piawai. Dengan membandingkan antara rumusan rambatan balik piawai dan algoritma Krzyzak, didapati bahawa $(1 - O_{pk}) O_{pk}$ telah dihapuskan, oleh itu masalah minimum setempat (*local minimum*) boleh diatasi.

UJIKAJI TERHADAP ALGORITMA RAMBATAN BALIK PIAWAI DAN ALGORITMA KRZYZAK

Pengujian terhadap rambatan balik piawai (BP) dan algoritma Krzyzak dilakukan dengan menggunakan data data siri masa bagi tujuan peramalan iaitu data bulanan bagi pengeluaran kayu balak negara dari tahun 1981 hingga tahun 1994 (Lampiran 1). Data tersebut terlebih dahulu ditukarkan ke bentuk $[0,1]$ dengan menggunakan jelmaan linear, iaitu

$$X_n = \frac{X_0 - X_{min}}{X_{max} - X_{min}}$$

dengan X_n nilai yang telah dinormalkan, X_0 nilai data asal, X_{min} nilai minimum data dan X_{max} nilai maksimum data.

Model rangkaian neural dibangunkan bagi meramalkan pengeluaran pada satu bulan, enam bulan dan duabelas bulan ke hadapan iaitu mewakili ramalan jangka pendek, sederhana dan jangka panjang. Bilangan nod input ditetapkan kepada 12 nod, manakala bilangan nod output pula berdasarkan bilangan bulan yang ingin diramalkan iaitu 1, 6 dan 12. Beberapa struktur rangkaian dengan bilangan nod tersembunyi 4, 8, 12 dan 24 serta nilai pemalar pembelajaran $\alpha=0.1$, 0.5 dan 0.9 dan pemalar momentum $\beta=0.5$ digunakan. Ralat punca kuasa dua (RMS) dikira bagi menentukan ketepatan model rangkaian.

Hasil Kajian

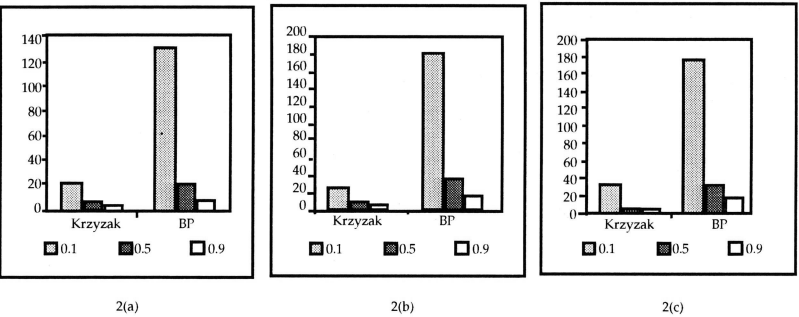
Hasil ujikaji dibahagikan kepada tiga bahagian iaitu (1) Kesan Pemalar Pembelajaran, α (2) Kesan Bilangan Nod Tersembunyi, N dan (3) Kesan Bilangan Nod Output.

Kesan Pemalar Pembelajaran (α)

Jadual 1 menunjukkan masa pembelajaran bagi peramalan jangka pendek (1 bulan ke hadapan), iaitu untuk 1 nod output. Nilai momen-tum yang digunakan adalah 0.5. Apabila pemalar pembelajaran ditingkatkan, diperoleh masa pembelajaran bagi kedua algoritma dapat dikurangkan dan algoritma Krzyzak menunjukkan prestasi yang lebih baik berbanding algoritma BP (Rajah 2).

Jadual 1
Masa Pembelajaran Algoritma Krzyzak dan BP dengan Struktur 12:N:1 Menggunakan Beberapa Nilai Pemalar Pembelajaran

N	4			12			24		
α	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
Krzyzak	35	6	5	28	4	2	24	3	1
BP	178	32	19	181	36	19	131	22	11



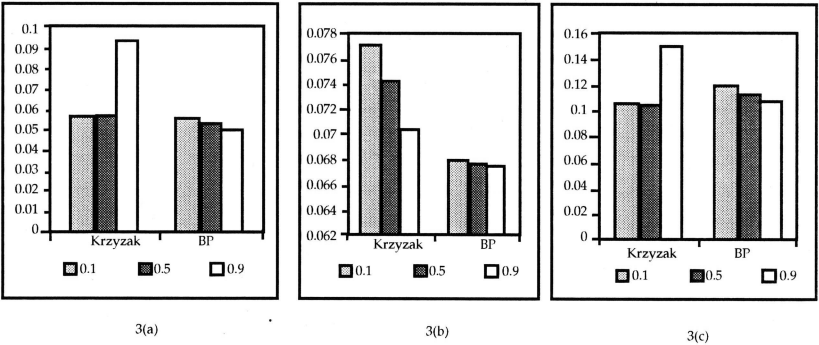
Rajah 2
Masa Pembelajaran (a) nod tersembunyi 4 (b) nod tersembunyi 12 (c) nod tersembunyi 24

Dari segi ketepatan pula (Jadual 2), nilai pemalar pembelajaran yang kecil ($\alpha \leq 0.5$) adalah lebih baik bagi algoritma Krzyzak sekiranya bilangan nod tersembunyi adalah besar (>12). Sebaliknya, jika bilangan

nod tersembunyi adalah kecil maka, peningkatan nilai pemalar pembelajaran dapat mengurangkan nilai RMS. Pemalar pembelajaran tidak begitu mempengaruhi ketepatan algoritma BP. Algoritma BP adalah lebih tepat berbanding algoritma Krzyzak sekiranya bilangan nod tersembunyi adalah kecil (≤ 12). Penggunaan pemalar pembelajaran yang besar lebih sesuai untuk algoritma BP kerana ia bukan sahaja dapat meningkatkan masa pembelajaran malah nilai RMS juga semakin kecil. Rajah 3 menunjukkan perbandingan nilai RMS bagi kedua algoritma dengan bilangan nod tersembunyi 4, 12 dan 24.

Jadual 2
RMS Algoritma Krzyzak dan BP dengan Struktur 12:N:1
menggunakan Beberapa Nilai Pemalar Pembelajaran

N	4			12			24		
α	0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
Krzyzak	0.0771	0.0742	0.0705	0.0577	0.0582	0.0951	0.1046	0.1043	0.1494
BP	0.068	0.0678	0.0676	0.0575	0.0544	0.0514	0.1187	0.1128	0.1074



Rajah 3
RMS (a) nod tersembunyi 4 (b) nod tersembunyi 12
(c) nod tersembunyi 24

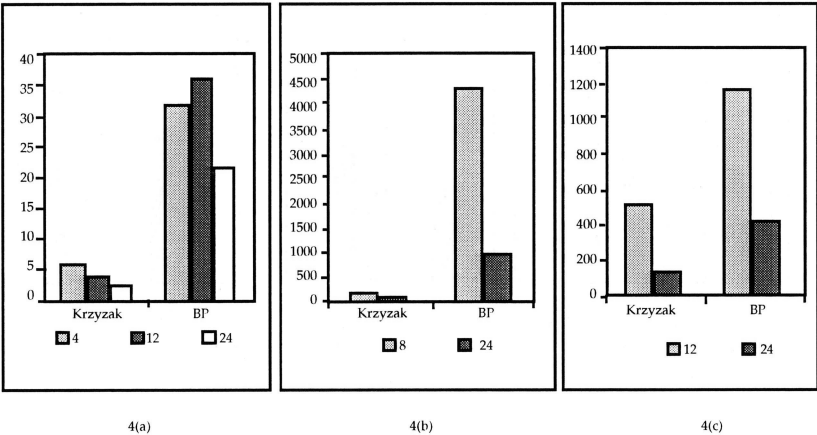
Kesan Bilangan Nod Tersembunyi

Jadual 3 menunjukkan masa pembelajaran kedua algoritma bagi peramalan jangka pendek, sederhana dan jangka panjang dengan beberapa bilangan nod tersembunyi. Nilai pemalar pembelajaran, $\alpha=0.5$ digunakan semasa larian. Didapati, masa pembelajaran adalah kecil apabila bilangan nod tersembunyi lebih besar, dan kedua-dua algoritma menunjukkan keadaan yang sama (Rajah 4). Namun begitu,

algoritma Krzyzak memberikan masa pembelajaran yang lebih baik berbanding BP.

Jadual 3
Masa Pembelajaran Algoritma Krzyzak dan BP dengan $\alpha=0.5$ menggunakan Beberapa Nod Tersembunyi dan Nod Output

Nod Output	1			6		12	
N	4	12	24	8	24	12	24
Krzyzak	6	4	3	132	74	530	157
BP	32	36	22	4349	1008	1172	432



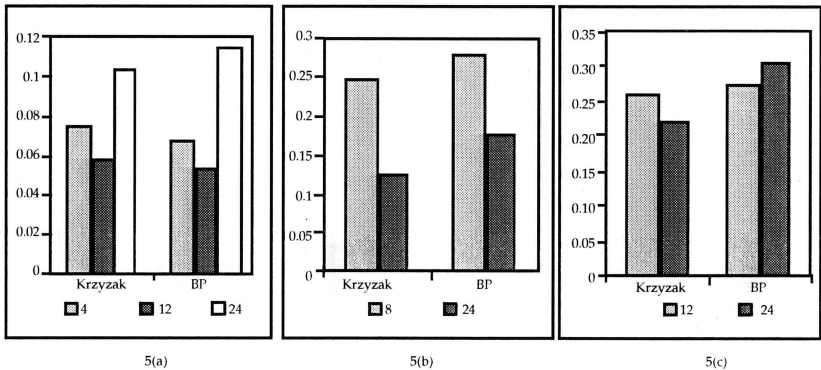
Rajah 4
Masa Pembelajaran (a) 1 nod output (b) 6 nod output (c) 12 nod output

Dari segi ketepatan pula (Jadual 4), bagi 1 nod output, kedua algoritma menunjukkan perlakuan yang sama. Algoritma BP lebih jitu berbanding algoritma Krzyzak apabila bilangan nod tersembunyi adalah kecil (≤ 12). Kedua algoritma juga menunjukkan perlakuan yang sama bagi 6 nod output, namun begitu algoritma Krzyzak memberikan nilai RMS yang lebih kecil. Bagi 12 nod output, kedua algoritma memberikan perlakuan yang berbeza. Algoritma Krzyzak bertambah jitu apabila bilangan nod tersembunyi besar, tetapi keadaan sebaliknya berlaku dalam algoritma BP. Algoritma Krzyzak memberikan RMS yang lebih baik apabila bilangan nod tersembunyi adalah besar (Rajah 5).

Jadual 4

RMS Algoritma Krzyzak dan BP dengan $\alpha=0.5$ menggunakan Beberapa Nod Tersembunyi dan Nod Output

Nod Output	1			6		12	
N	4	12	24	8	24	12	24
Krzyzak	0.0742	0.0582	0.1043	0.247954	0.124378	0.259438	0.221759
BP	0.0678	0.0544	0.1128	0.279346	0.179994	0.277418	0.304846



Rajah 5

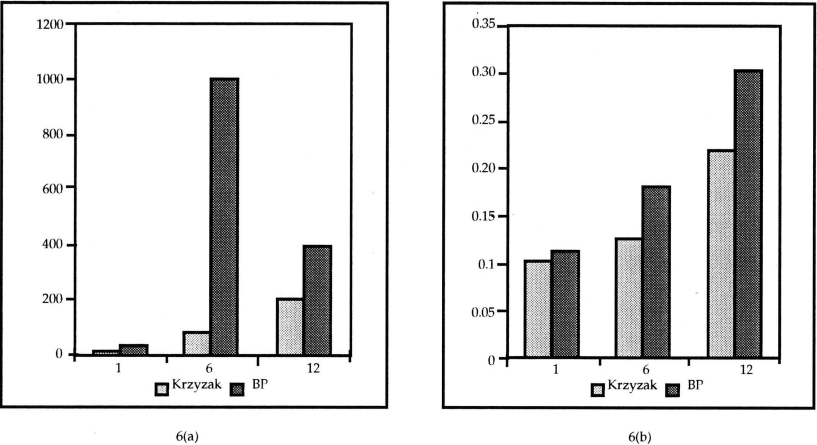
RMS (a) 1 nod output (b) 6 nod output (c) 12 nod output

Kesan Bilangan Nod Output

Jadual 5 menunjukkan masa pembelajaran dan RMS bagi kedua algoritma. Pemalar pembelajaran adalah 0.5, bilangan nod tersembunyi ialah 24 dengan bilangan nod output 1, 6 dan 12. Berdasarkan Rajah 6(a) didapati masa pembelajaran bagi algoritma Krzyzak meningkat apabila bilangan nod output bertambah. Bagi algoritma BP pula, masa yang agak lama diperlukan apabila bilangan nod output ialah 6. Dari segi ketepatan pula (Rajah 6(b)), kedua algoritma menunjukkan kesan yang sama, iaitu semakin besar bilangan nod output semakin besar nilai RMS. Bagi kesemua bilangan nod output, algoritma Krzyzak menunjukkan prestasi yang lebih baik berbanding algoritma BP baik dari segi masa mahupun ketepatan.

Jadual 5
Masa Pembelajaran dan RMS untuk Beberapa Nod Output
dengan $\alpha=0.5$ dan $N=24$

Nod Output	Masa			RMS		
	1	6	12	1	6	12
Krzyzak	3	74	157	0.1043	0.124378	0.221759
BP	22	1008	432	0.1128	0.179994	0.304846



Rajah 6
Algoritma Krzyzak dan BP dengan Beberapa Nod Output
(a) Masa Pembelajaran (b) RMS

KESIMPULAN

Algoritma Krzyzak sesuai digunakan dengan nilai pemalar pembelajaran yang kecil bagi mendapatkan ralat peramalan yang lebih kecil dengan masa pembelajaran yang pantas. Algoritma BP pula, memerlukan nilai pemalar pembelajaran yang lebih besar bagi mempercepatkan masa pembelajaran dan mengurangkan ralat peramalan. Prestasi algoritma Krzyzak lebih baik berbanding algoritma BP bagi peramalan jangka sederhana dan jangka panjang dengan penggunaan bilangan nod tersembunyi yang besar. Sebaliknya, algoritma BP adalah lebih baik bagi peramalan jangka pendek dengan bilangan nod tersembunyi yang kecil. Secara keseluruhannya dapat dirumuskan, algoritma Krzyzak menunjukkan prestasi yang lebih baik berbanding algoritma BP terutamanya dari segi masa pembelajaran.

RUJUKAN

- Fausett, L.V. (1994). *Fundamentals of neural networks: Architectures, algorithms and applications*. New Jersey: Prentice Hall.
- Siti Mariyam Hj. Shamsuddin, Norazah Yusof, Paridah Samsuri, Nor Bahiah Ahmad, & Siti Zaiton Mohd Hashim (2000, Disember). Pengkelasan tahap pengetahuan pelajar secara pembelajaran adaptiv berteknologikan hipermedia – satu pendekatan rangkaian neural. *Jurnal Teknologi Maklumat*, 12(2), 1-10.
- Krzyzak, W. & Shen, C.Y. (1990). Classification of large set of hand-written characters using modified BP model. *Proceedings of the International Joint Conference on Neural Networks* (p.225-232). San Diego.
- Timothy, M. (1993). *Practical Neural Network Recipes in C++*. San Diego, CA: Academic Press.
- Roselina Sallehuddin, Mohd Salihin Ngadiman, & Siti Mariyam Hj Shamsuddin (2001). Peramalan siri masa bermusim menggunakan rangkaian neural terhadap nyahmusim data. *Sains Malaysiana*, 30, 107-118.

Pengeluaran Bulanan Kayu Balak Negara dari Tahun 1981 hingga 1994

LAMPIRAN 1

Tahun	Januari	Februari	Mac	April	Mei	Jun	Julai	Ogos	September	Oktober	November	Disember
1981	45399	45097	50085	54322	84960	60209	48291	82370	68250	82406	96757	104211
1982	49088	91035	83671	96282	110152	115379	74535	95723	119309	111220	158826	157697
1983	61098	108643	85600	61462	58408	88939	131824	139453	72232	104155	75750	106235
1984	59198	54268	104977	84544	89816	75334	57871	124326	132265	47396	123537	306549
1985	60676	121552	59616	149154	124880	130836	139856	139343	94670	111549	114813	156564
1986	66432	58192	139714	45954	82167	110094	66534	152398	120937	97728	104895	245778
1987	50827	160354	158363	127456	193404	139828	143055	201403	166570	224314	188662	151706
1988	116314	104379	132667	118855	149606	204744	97671	211483	197846	128906	205130	182105
1989	106398	161260	212097	155016	206241	336493	201421	266533	304921	261595	246115	212439
1990	218192	163936	246209	263497	236277	274689	234672	245925	230651	254709	281976	232159
1991	263957	181203	266759	251298	241285	227667	289976	310447	247809	317902	285747	259469
1992	262556	147918	247539	303610	288117	274564	273023	183684	282951	247531	283628	376349
1993	192232	200757	325474	343272	283097	260278	214687	236178	240113	213118	180437	178414
1994	212780	133835	215372	222030	222811	202603	242368	220399	260867	192345	235851	207542

Sumber: Siaran Perangkaan Bulanan Sarawak (1981-1995). Jabatan Perangkaan Malaysia (Cawangan Sarawak), 1981-1995.

